

KAIST X ESTSOFT 인턴십

송건우 홍민희 고어진



Table of Contents

#1 문제 정의

문제에서 고려한 사항 및 조건

#2 베이스 라인 모델

베이스 라인 모델 학습 방법 및 결과

#3 지표 제안

지표 정의
제안한 이유

#4 모델 개선 아이디어

지표를 포함하도록 모델을 개선

#5 아이디어 평가

개선한 모델 결과

#6 결론 및 피드백

제안한 지표에 대한 평가
피드백

#01문제 정의

#안정성

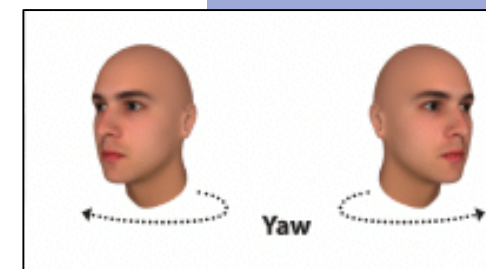
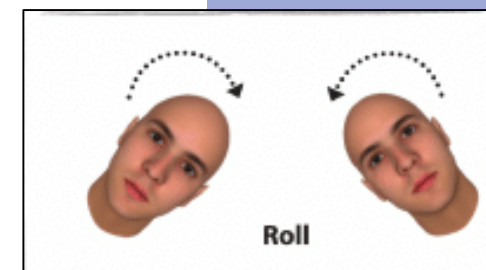
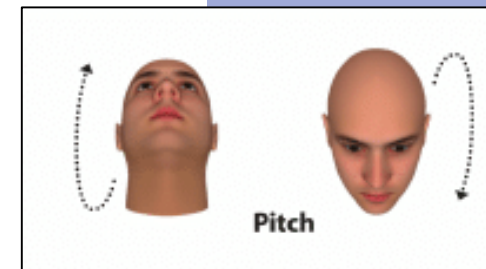
연속된 프레임에서 개별 키 포인트의 얼굴 위에서 상대적 위치 변화가 오차범위 안에 존재하면 안정적이라고 판단

#고려 범위

1280 x 720 사이즈의 영상만 고려함

연속한 20개의 random frame을 모델 학습에 사용

장면 전환은 고려하지 않음



SUMMARY

Layer (type)	Output Shape	Param #
Conv2d-1	[-1, 16, 1280, 720]	1,216
BatchNorm2d-2	[-1, 16, 1280, 720]	32
ReLU-3	[-1, 16, 1280, 720]	0
MaxPool2d-4	[-1, 16, 320, 240]	0
Dropout-5	[-1, 16, 320, 240]	0
Conv2d-6	[-1, 32, 320, 240]	12,832
ReLU-7	[-1, 32, 320, 240]	0
Conv2d-8	[-1, 32, 320, 240]	9,248
BatchNorm2d-9	[-1, 32, 320, 240]	64
ReLU-10	[-1, 32, 320, 240]	0
MaxPool2d-11	[-1, 32, 80, 80]	0
Dropout-12	[-1, 32, 80, 80]	0
Conv2d-13	[-1, 64, 80, 80]	18,496
ReLU-14	[-1, 64, 80, 80]	0
Conv2d-15	[-1, 64, 80, 80]	36,928
BatchNorm2d-16	[-1, 64, 80, 80]	128
ReLU-17	[-1, 64, 80, 80]	0
MaxPool2d-18	[-1, 64, 40, 40]	0
Dropout-19	[-1, 64, 40, 40]	0
Conv2d-20	[-1, 128, 40, 40]	73,856
ReLU-21	[-1, 128, 40, 40]	0
Conv2d-22	[-1, 128, 40, 40]	147,584
BatchNorm2d-23	[-1, 128, 40, 40]	256
ReLU-24	[-1, 128, 40, 40]	0
MaxPool2d-25	[-1, 128, 20, 20]	0
Dropout-26	[-1, 128, 20, 20]	0
Conv2d-27	[-1, 256, 20, 20]	295,168
ReLU-28	[-1, 256, 20, 20]	0
Conv2d-29	[-1, 256, 20, 20]	590,080
BatchNorm2d-30	[-1, 256, 20, 20]	512
ReLU-31	[-1, 256, 20, 20]	0
MaxPool2d-32	[-1, 256, 10, 10]	0
Dropout-33	[-1, 256, 10, 10]	0
Conv2d-34	[-1, 512, 10, 10]	1,180,160
ReLU-35	[-1, 512, 10, 10]	0
Conv2d-36	[-1, 512, 10, 10]	2,359,608
BatchNorm2d-37	[-1, 512, 10, 10]	1,024
ReLU-38	[-1, 512, 10, 10]	0
Dropout-39	[-1, 512, 10, 10]	0
MaxPool2d-40	[-1, 512, 5, 5]	0
Linear-41	[-1, 512]	6,554,112
ReLU-42	[-1, 512]	0
Linear-43	[-1, 136]	69,768

Total params: 11,351,272
 Trainable params: 11,351,272
 Non-trainable params: 0

Input size (MB): 10.55
 Forward/backward pass size (MB): 485.65
 Params size (MB): 43.30
 Estimated Total Size (MB): 539.50

#02

베이스 라인 모델

#03지표 제안

JITTER

연속된 프레임에 대해
점이 흔들리는 현상

비지도학습

DBSCAN, PCA 등의
비지도학습을 이용해
각 키 포인트들 군집화

VECTOR

얼굴의 3차원 회전,
이동을 고려한 벡터를
지표로 사용

주어진 Ground Truth를 이용한 지표

#1 통계학적 지표

연속한 프레임에서
키 포인트의 변위가
정규분포를 따른다고
가정

#2 편차 이용

눈 사이 간격과 코 길이를
각각 x 방향 기준 값과
y 방향 기준 값으로 생각

$$\left(\frac{\Delta x_t - \overline{\Delta x}}{|P_{37} P_{46}|}, \frac{\Delta y_t - \overline{\Delta y}}{|P_{28} P_{34}|} \right)$$

#3 길이 비 이용

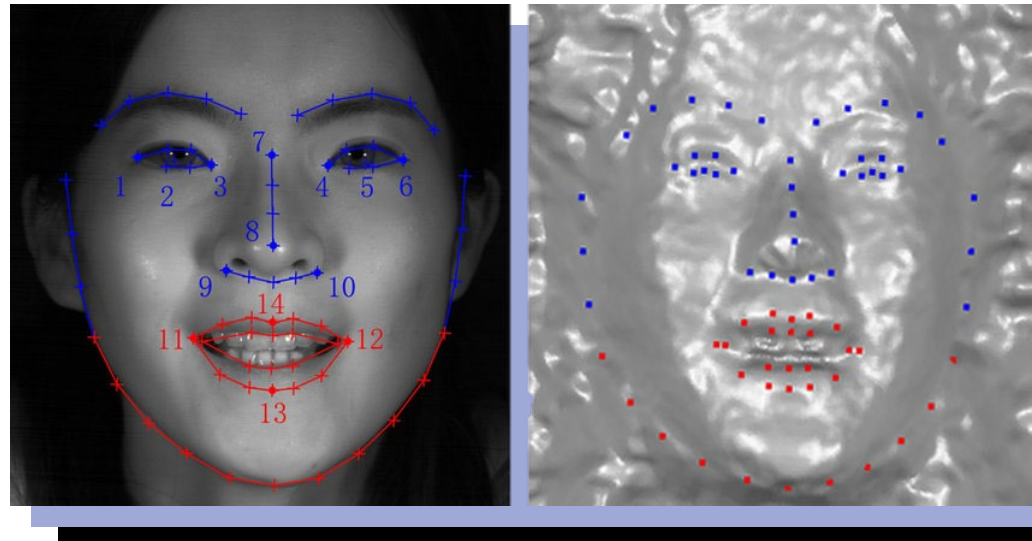
#2와 같은 기준 값으로
편차와 기준 값의 길이 비가
일정할 때 안정하다고 생각

$$\frac{x_k - \overline{x_k}}{x_{k-1} - \overline{x_{k-1}}} \cdot \frac{l_{k-1}}{l_k}$$

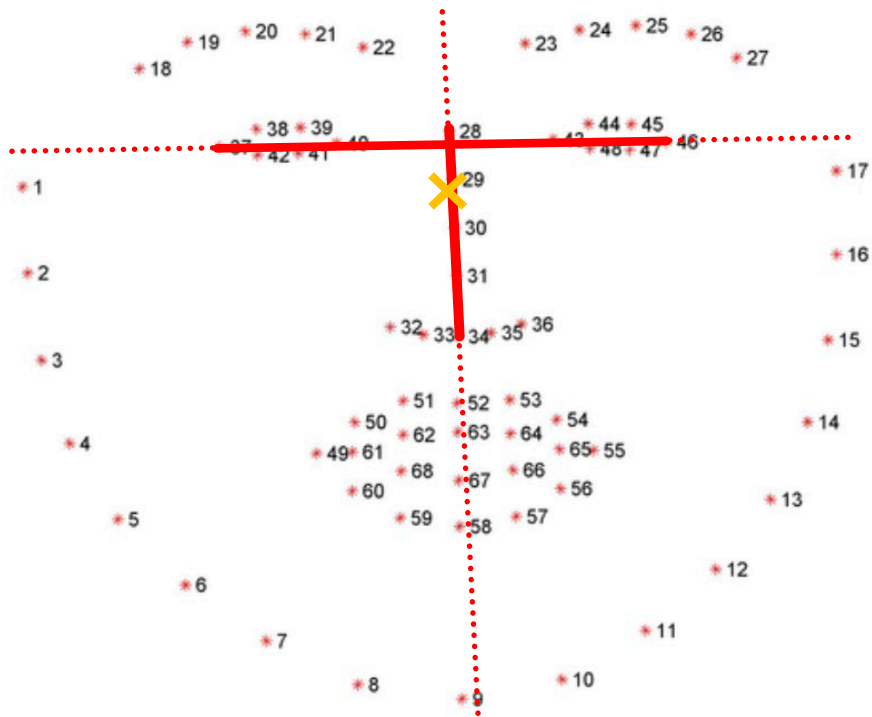
최종 지표

#가정

얼굴에서 양쪽 눈꼬리 점과
코 위아래 끝 점의 위치가
고정되어 있다고 가정함



- Blue: static points
- Red: dynamic points



$P_{37}, P_{46}, P_{28}, P_{34}$ 의

scaling

/ ≈ 1

#04 모델 개선 아이디어

#Loss function 보정

Loss function이 앞서 제안한 지표를 포함하도록 보정

```
if prev_loss < 100:  
    if count % continuous == 0:  
        prev_out_calcs = None  
        result = out.clone().detach()  
        prev_out_calcs, indicator_result = indicator(result, prev_out_calcs)  
        loss = criterion(out, target) + indicator_result #mse + indicator
```

160 Epoch loss: 98.9207023549963
 161 Epoch loss: 98.2147507749222
 162 Epoch loss: 108.49963011388425
 163 Epoch loss: 111.57241899216616
 164 Epoch loss: 101.0350798037317
 165 Epoch loss: 101.93312264548408
 166 Epoch loss: 101.89911405532449
 167 Epoch loss: 96.45696873885613
 168 Epoch loss: 104.34381597969266
 169 Epoch loss: 108.52387872426598
 170 Epoch loss: 120.64322653059607
 171 Epoch loss: 161.1126399450832
 172 Epoch loss: 151.78984360739037
 173 Epoch loss: 149.4278505175202
 174 Epoch loss: 115.52429145088902
 175 Epoch loss: 113.58428722443404
 176 Epoch loss: 98.67745889866794
 177 Epoch loss: 113.48312989009752
 178 Epoch loss: 128.7250592885194
 179 Epoch loss: 125.41415312312267
 180 Epoch loss: 116.54287593408867
 181 Epoch loss: 99.47491272687913
 182 Epoch loss: 96.8712107177134
 183 Epoch loss: 115.8604423628913
 184 Epoch loss: 98.818521870176
 185 Epoch loss: 108.40973360119042
 186 Epoch loss: 93.67631509789715
 187 Epoch loss: 103.38919907190181
 188 Epoch loss: 115.55302032795217
 189 Epoch loss: 132.45641549781516
 190 Epoch loss: 100.08234173059464
 191 Epoch loss: 97.03996796431365
 192 Epoch loss: 85.07457058319339
 193 Epoch loss: 91.30055846660225
 194 Epoch loss: 107.23015652718368
 195 Epoch loss: 108.53615979684724
 196 Epoch loss: 109.07113648123212
 197 Epoch loss: 88.34820878373252
 198 Epoch loss: 92.4337045263361
 199 Epoch loss: 100.98190804675774

160 Epoch loss: 103.12510738726016
 161 Epoch loss: 104.21704645068557
 162 Epoch loss: 93.4022702049326
 163 Epoch loss: 100.11811539464527
 164 Epoch loss: 92.14004717402987
 165 Epoch loss: 99.70087076955372
 166 Epoch loss: 100.33697561378833
 167 Epoch loss: 86.5809152159426
 168 Epoch loss: 100.64055986271964
 169 Epoch loss: 124.24925304761639
 170 Epoch loss: 115.45624470512072
 171 Epoch loss: 116.70989465978411
 172 Epoch loss: 97.38964082709064
 173 Epoch loss: 104.65093486309051
 174 Epoch loss: 102.14299673592603
 175 Epoch loss: 94.73009574060087
 176 Epoch loss: 92.92223213712374
 177 Epoch loss: 93.0044631185355
 178 Epoch loss: 106.90424549358862
 179 Epoch loss: 111.17975949932028
 180 Epoch loss: 109.18504473500782
 181 Epoch loss: 104.44084026769356
 182 Epoch loss: 96.75743010176552
 183 Epoch loss: 107.75529245049866
 184 Epoch loss: 101.40872009860145
 185 Epoch loss: 91.12348635108383
 186 Epoch loss: 101.51270607400824
 187 Epoch loss: 99.19525372496358
 188 Epoch loss: 97.30161460152378
 189 Epoch loss: 98.05506602304953
 190 Epoch loss: 112.51540398818476
 191 Epoch loss: 117.35602449487757
 192 Epoch loss: 113.93036272768622
 193 Epoch loss: 101.27262895018966
 194 Epoch loss: 93.31165861465313
 195 Epoch loss: 98.07441765092037
 196 Epoch loss: 110.76551714782362
 197 Epoch loss: 98.83654982867064
 198 Epoch loss: 104.09001537208204
 199 Epoch loss: 93.4665808684296

모델 트레이닝 로스

좌: 베이스

우: 지표 포함

Model MSE: 76438.208672

Model with Indicator MSE: 84739.324531

비디오 548 MSE 결과

상: 베이스

하: 지표 포함

Indicator value: 1.005329365435061

Indicator value: 1.002679322512951

비디오 548 지표 결과

상: 베이스

하: 지표 포함



KAIST X ESTSOFT
INTERNSHIP

#05

아이디어 평가

#06 결론 및 피드백

- 주어진 Ground Truth에서는 새로운 지표의
- 하지만 트레이닝 도중에는 연속적인 프레임의 차이가 클 수도 있다
- 지표를 더 정확하게 모델 학습 중에 쓰려면 학습이 잘 된 상태에서 해야 한다 (Low Loss)
- 지표 모델의 성능 저하는 너무 가벼운 모델에서 학습이 덜 된 상태에서 써서 그렇다



- 더 정확한 모델로 더 Rigorous한 학습으로 지표의 효율을 확인 해야 한다

빨강: Ground Truth
초록: 베이스라인
파랑: 지표 포함

감사합니다

KAIST X ESTSOFT INTERNSHIP

A조 송건우 홍민희 고어진